

PH NL 021175	MAT. DOSSIER
-----------------	-----------------

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/31418 A2

- (51) International Patent Classification⁷: G06F
- (21) International Application Number: PCT/US00/41571
- (22) International Filing Date: 25 October 2000 (25.10.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/161,587 26 October 1999 (26.10.1999) US
- (71) Applicant: ARTHUR D. LITTLE, INC. [US/US]; Acorn Park, Cambridge, MA 02140 (US).
- (72) Inventors: JACKSON, James, H.; 119 Penwood Drive, Cary, NC 27511 (US); KRAUS, Thomas, D.; 480 Walnut Street, Newtonville, MA 02460 (US).
- (74) Agents: LEBOVICI, Victor, B. et al.; Weingarten, Schurigin, Gagnebin & Hayes, LLP, Ten Post Office Square, Boston, MA 02109 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/31418 A2

(54) Title: WIDE CONNECTIONS FOR TRANSFERRING DATA BETWEEN PE'S OF AN N-DIMENSIONAL MESH-CONNECTED SIMD ARRAY WHILE TRANSFERRING OPERANDS FROM MEMORY

(57) Abstract: A single-instruction multiple-data (SIMD) array processor providing enhanced data transfer efficiency. The SIMD array processor includes at least one memory and a plurality of mesh-connected processing elements configured in an array. Each processing element in the array includes at least one "narrow" memory buffer, at least one "wide" data register, and at least one "wide" communication register. The narrow memory buffer is adapted to transfer data serially between the memory and the wide data register, the wide data register is adapted to transfer data directly to the wide communication register, and the wide communication register is adapted to transfer data directly to the communication register of a neighboring processing element while the memory buffer is accessing data from the memory.

-1-

TITLE OF THE INVENTION

WIDE CONNECTIONS FOR TRANSFERRING DATA BETWEEN PE'S OF AN
N-DIMENSIONAL MESH-CONNECTED SIMD ARRAY WHILE
5 TRANSFERRING OPERANDS FROM MEMORY

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority of U.S. Provisional
Patent Application No. 60/161,587 filed October 26, 1999
10 entitled FINITE DIFFERENCE ACCELERATOR.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR
DEVELOPMENT

N/A

15

BACKGROUND OF THE INVENTION

The present invention relates generally to SIMD
array processors, and more specifically to SIMD array
processors having improved data transfer efficiency
20 between processing elements incorporated therein.

Single-Instruction Multiple-Data (SIMD) array
processors are known which comprise multi-dimensional
arrays of interconnected processing elements executing
the same instruction simultaneously on a plurality of
25 different data samples. For example, an SIMD array
processor may include a two-dimensional array of
processing elements in which each processing element is
connected to its four (4) nearest neighboring processing
elements to form a "North, East, West, South (NEWS)
30 array". In such NEWS arrays, each processing element can

-2-

communicate directly with its North, East, West, and South neighbors.

One aspect of the typical SIMD array processor that limits the rate at which the processing elements can communicate with each other is that one or more of the neighboring processing elements with which a particular processing element communicates may be physically located on a different Application Specific Integrated Circuit (ASIC) and/or on a different Printed Circuit Board (PCB). For example, when a processing element directly accesses a multi-bit data sample from a neighboring processing element physically located on a different ASIC or a different PCB, a significant amount of time may be required for that data sample to propagate between the ASIC's or PCB's. To account for this propagation time, communication registers used in processing elements of the typical SIMD array processor are generally clocked at relatively low speeds. However, clocking communication registers at such low speeds may cause many operating cycles of a processing element to be wasted while the processing element waits for the data transfer to complete. As a result, the typical SIMD array processor may not be suitable for some high-speed data processing applications.

It would therefore be desirable to have an SIMD array processor that has improved data transfer efficiency between processing elements incorporated therein. Such an SIMD array processor would transfer data samples more efficiently whether or not neighboring processing elements are located on the same ASIC or PCB.

-3-

BRIEF SUMMARY OF THE INVENTION

In accordance with the present invention, an SIMD array processor is provided in which data transfer efficiency is enhanced between processing elements included therein. The SIMD array processor includes a plurality of mesh-connected processing elements configured in a multi-dimensional array. Each processing element includes at least one "narrow" memory buffer, at least one "wide" data register, and at least one "wide" communication register. The narrow memory buffer is adapted to transfer data serially between a memory and the wide data register; and, the wide data register is adapted to transfer data directly to the wide communication register. Further, the wide communication register is adapted to transfer data directly to the communication register of a neighboring processing element while the memory buffer accesses data from memory. In a preferred embodiment, the memory buffer has a width of one (1) bit to allow bit-serial data transfer between the memory and the wide data register while the wide communication register transfers data in parallel to (from) the communication register of the neighboring processing element.

Other features, functions, and aspects of the invention will be evident from the Detailed Description of the Invention that follows.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

The invention will be more fully understood with reference to the following Detailed Description of the Invention in conjunction with the drawings of which:

-4-

Fig. 1 is a block diagram depicting an SIMD array processor according to the present invention;

Fig. 2 is a block diagram depicting a processing element included in the SIMD array processor of Fig. 1; and

Fig. 3 is a table depicting the contents of respective YS, NEWSO, and NEWSI registers of two (2) neighboring processing elements included in the SIMD array processor of Fig. 1, in an illustrative example of a data transfer between the two (2) processing elements.

DETAILED DESCRIPTION OF THE INVENTION

The entire disclosure of U.S. Provisional Patent Application No. 60/161,587 filed October 26, 1999 is incorporated herein by reference.

Fig. 1 is a block diagram depicting an illustrative embodiment of an SIMD array processor 100 in accordance with the present invention. The SIMD array processor 100 includes a plurality of identical Processing Elements (PE's) 104 through 134 interconnected in a mesh and configured as a two-dimensional NEWS array. The PE's 104 through 134 are communicably coupled to respective bi-directional Input/Output's (I/O's) 0 through 15 of a Random Access Memory (RAM) 102. In a preferred embodiment, the RAM 102 is a Synchronous Dynamic RAM (SDRAM).

Although Fig. 1 depicts the SIMD array processor 100 as including the two-dimensional NEWS array of PE's 104 through 134, it should be understood that the SIMD array processor 100 may comprise an array of PE's having one (1) or more dimensions. It should also be understood

-5-

that the size of the array may be adjusted. In a preferred embodiment, the NEWS array of PE's is implemented on an ASIC comprising an 8 x 8 NEWS array. Fig. 1 depicts the SIMD array processor 100 as including
5 a 4 x 4 NEWS array for clarity of discussion.

In a preferred embodiment, the PE's 104 through 134 are physically located on the same ASIC to form the NEWS array, thereby simplifying I/O connections between nearest neighboring PE's in the NEWS array. For example,
10 relative to the PE 114, a "North" I/O 140 interconnects the PE 114 and the PE 112; an "East" I/O 144 interconnects the PE 114 and the PE 122; a "West" I/O 146 interconnects the PE 114 and the PE 106; and, a "South" I/O 142 interconnects the PE 114 and the PE 116.
15 Further, the PE's 104, 106, 108, 110, 112, 118, 120, 126, 128, 130, 132, and 134 that are physically located along the edges of the NEWS array comprise suitable North, East, West, and South I/O's for connecting these PE's with PE's that are physically located on different
20 ASIC's. For example, a 2 x 2 array of ASIC's may be implemented on a PCB; and, North, East, West, and South I/O's of PE's physically located along the edges of NEWS arrays of adjacent ASIC's may be suitably interconnected. Still further, respective 2 x 2 arrays of ASIC's may be
25 implemented on different PCB's; and, the North, East, West, and South I/O's of PE's physically located along the edges of the respective arrays of ASIC's may be suitably interconnected. In this way, the size of the NEWS array comprising the plurality of identical PE's can
30 be increased to satisfy the processing requirements of the target application.

-6-

In this illustrative embodiment, each of the PE's 104 through 134 is connected to one and only one of the I/O's 0 through 15 of the RAM 102. For example, the PE's 104, 106, 108, and 110 are connected to the I/O's 15, 14, 13, and 12, respectively; the PE's 112, 114, 116, and 118 are connected to the I/O's 11, 10, 9, and 8, respectively; the PE's 120, 122, 124, and 126 are connected to the I/O's 7, 6, 5, and 4, respectively; and, the PE's 128, 130, 132, and 134 are connected to the I/O's 3, 2, 1, and 0, respectively, of the RAM 102. In a preferred embodiment, the PE's 104 through 134 utilize the I/O's 0 through 15 of the RAM 102 to access data in the RAM 102 in a bit-serial fashion.

Those of ordinary skill in this art will appreciate that the SIMD array processor 100 of Fig. 1 may be incorporated in a multi-dimensional processing system including, e.g., a command preprocessor interfaced with a processor controller that provides intermediary processing functions between the command preprocessor and the SIMD array processor 100.

Fig. 2 is a block diagram depicting an illustrative embodiment of a PE 200 in accordance with the present invention. In a preferred embodiment, the PE 200 is representative of each of the PE's 104 through 134 included in the SIMD array processor 100. Accordingly, the SIMD array processor 100 preferably includes a plurality of identical PE's such as the representative PE 200 interconnected in the 4 x 4 NEWS array configuration.

In the illustrated embodiment, the PE 200 includes a multi-bit data register YS 204. In a preferred embodiment, the data register YS 204 is 64-bits wide. It

-7-

should be appreciated that the data register YS 204 can be used to store, e.g., a floating-point number or a signed/unsigned fixed-point integer. It should also be appreciated that the data register YS 204 can be used as
5 a shift register. For example, the data register YS 204 may be used to shift in binary values in a bit-serial fashion from a memory buffer M 202 by way of a line 216. In a preferred embodiment, the memory buffer M 202 is a 1-bit wide memory buffer. Moreover, the data register YS
10 204 may be used to shift left binary values contained therein.

As described above, the PE's 104 through 134 (see Fig. 1) are preferably physically located on an ASIC to form a NEWS array to simplify I/O interconnections
15 between the nearest neighboring PE's in the NEWS array. In a preferred embodiment, each of the PE's 104 through 134 reads four (4) bits of data directly (in parallel) from a nearest neighboring PE connected to its North, East, West, or South I/O while writing four (4) bits of
20 data in parallel to an opposite neighboring PE. For example, the PE 114 may read four (4) bits of data in parallel from the PE 112 via the North I/O 140 while writing four (4) bits of data in parallel to the opposite neighbor PE 116 via the South I/O 142, and vice versa
25 (see Fig. 1); and, the PE 114 may read four (4) bits of data in parallel from the PE 106 via the West I/O 146 while writing four (4) bits of data in parallel to the opposite neighbor PE 122 via the East I/O 144, and vice versa (see Fig. 1).

30 In the illustrated embodiment, the PE 200 (see Fig. 2) includes communication registers NEWS Input (NEWSI)

-8-

206 and NEWS Output (NEWSO) 208. In a preferred embodiment, each of the communication registers NEWSI 206 and NEWSO 208 is 4 bits wide for reading (writing) four (4) bits of data in parallel from (to) a nearest neighboring PE connected to its North, East, West, or South I/O. Specifically, the communication register NEWSI 206 receives four (4) data bits in parallel from one of the nearest neighboring North, East, West, and South PE's and preferably loads the four (4) data bits in parallel into the four (4) Least Significant Bit (LSB) positions (i.e., bit positions 0, 1, 2, and 3) of the data register YS 204 by way of a bus 218. Further, the data register YS 204 preferably loads four (4) data bits in parallel into the communication register NEWSO 208 by way of a bus 214 from its four (4) LSB positions for subsequent provision in parallel to one of the nearest neighboring North, East, West, and South PE's. The PE 200 includes a multiplexor (MUX) 210 that selects one of the 4-bit wide data buses coupled to respective communication registers NEWSO (not shown) of the nearest neighboring North, East, West, or South PE. The PE 200 also includes circuitry that enables the output of the communication register NEWSO 208 to drive only the bus that is connected to the nearest PE in the opposite direction. It is noted that the SIMD array processor 100 preferably includes a sequencer (not shown) for, e.g., decoding commands provided by a processor controller to obtain a stream of instructions and broadcasting the instruction stream to the NEWS array of PE's 104 through 134. Such a sequencer includes a control register for providing a 2-bit control word R that controls the

-9-

selection of one of the above-mentioned four (4) data buses by the MUX 210 and the bus that the communication register NEWSO 208 drives with its output.

5 The PE 200 includes the memory buffer M 202, which preferably comprises a 1-bit wide memory buffer through which all data are transferred between the PE 200 and the RAM 102. In a preferred embodiment, the memory buffer M 202 included in each of the PE's 104 through 134 is connected to one and only one of the I/O's 0 through 15
10 of the RAM 102 for accessing a single bit of data during each memory cycle of the RAM 102. For example, a single data bit may be read from one of the I/O's by way of the memory buffer M 202 and provided to the data register YS 204.

15 As explained above, the PE's 104 through 134 (see Fig. 1) are preferably physically located on the same ASIC to form a NEWS array, and arrays of such ASIC's may be implemented on respective PCB's with North, East, West, and South I/O's of nearest neighboring PE's
20 suitably interconnected. As a result, one or more of the nearest neighboring PE's with which a particular PE communicates may be physically located on a different ASIC and/or on a different PCB.

Those of ordinary skill in this art will appreciate
25 that when a PE directly accesses a multi-bit data sample from a neighboring PE physically located on a different ASIC or a different PCB, a significant amount of time may be required for that data sample to propagate between the ASIC's or PCB's. In a preferred embodiment, the SIMD
30 array processor 100 includes circuitry for generating a clock signal for latching four (4) bits of data in the

-10-

respective communication registers NEWSI 206 and NEWSO 208; and, each ASIC preferably includes circuitry for generating respective clock signals used by the PE's 104 through 134 and the RAM 102 physically located thereon.

5 The clock speed of the communication registers NEWSI 206 and NEWSO 208 is preferably one-fourth of the memory clock speed; and, the memory clock speed is preferably one-half of the PE clock speed. This means that the communication registers NEWSI 206 and NEWSO 208 can latch

10 their respective inputs once every eight (8) PE clock cycles. As a result, the communication register NEWSO 208 holds its latched data for nearly eight (8) PE clock cycles, and the communication register NEWSI 206 reads four (4) bits of data from the communication register

15 NEWSO of a neighboring PE (which may be physically located on a different ASIC or a different PCB) only after the data has settled for several consecutive PE clock cycles, thereby allowing sufficient time for the data to propagate.

20 The embodiments disclosed herein will be better understood with reference to the following illustrative example. In the following example, it is understood that the operating speed of the PE is twice that of the memory to allow the PE to perform two (2) sequential sets of

25 operations on registers included therein while the memory either provides a single bit to the memory buffer M or receives a single bit from the memory buffer M.

This example comprises the transfer of a 12-bit operand "A" (i.e., "A₁₁A₁₀A₉A₈A₇A₆A₅A₄A₃A₂A₁A₀ =

30 101001010011") from memory, to a memory buffer M0, to a data register YS0, and then to a communication register

-11-

NEWS00 of a first PE; and, the transfer of this operand from the communication register NEWS00 to a communication register NEWS11 and a data register YS1 of a second PE. In this example, it is understood that the first and second PE's are nearest neighboring PE's that may or may not be physically located on the same ASIC or the same PCB. It is further understood that the data registers YS0 and YS1 are 12-bit registers, the communication registers NEWS00 and NEWS11 are 4-bit registers, and the memory buffer M0 of the first PE is a 1-bit register. Further, at least the first PE includes circuitry configured for addressing a bit-serial column of memory. For example, binary values representative of the operand "A" may be stored in contiguous bit locations of the bit-serial column of memory. Moreover, the clock speed of the first and second PE's is understood to be twice that of the memory.

In this example, Fig. 3 depicts the contents of the YS0, NEWS00, NEWS11, and YS1 registers and the contents of the memory buffer M0 at the end of each memory cycle. During the first memory cycle, the memory provides the bit "A₁₁ = 1" (i.e., the Most Significant Bit (MSB) of the operand "A") to the memory buffer M0. Because the clock speed of the first PE is twice that of the memory, the bit "A₁₁ = 1" is transferred from the memory to the memory buffer M0 during the first half of the memory clock cycle, and the bit "A₁₁ = 1" is transferred from the memory buffer M0 to the LSB position of the data register YS0 during the second half of the memory clock cycle. Fig. 3 therefore depicts the contents of the memory buffer M0 as "1", and the respective contents of the YS0,

-12-

NEWS00, NEWS11, and YS1 registers as "XXXXXXXXXXXX1",
"XXXX", "XXXX", "XXXXXXXXXXXX" ("X" = "don't care") at
the end of the first memory cycle. During the second
memory cycle, the bit " $A_{10} = 0$ " is transferred from the
5 memory to the memory buffer M0 during the first half of
the memory clock cycle, and the first PE shifts the
contents of the data register YS0 left by one bit
position and loads the vacated bit position with the
contents of the memory buffer M0 during the second half
10 of the memory clock cycle. Fig. 3 therefore depicts the
contents of the memory buffer M0 as "0", and the
respective contents of the YS0, NEWS00, NEWS11, and YS1
registers as "XXXXXXXXXX10", "XXXX", "XXXX",
"XXXXXXXXXXXX" at the end of the second memory cycle.
15 During the third memory cycle, the bit " $A_9 = 1$ " is
transferred from the memory to the memory buffer M0
during the first half of the memory clock cycle, and the
first PE shifts the contents of the data register YS0
left by one bit position and loads the vacated bit
20 position with the contents of the memory buffer M0 during
the second half of the memory clock cycle. Fig. 3
therefore depicts the contents of the memory buffer M0 as
"1", and the respective contents of the YS0, NEWS00,
NEWS11, and YS1 registers as "XXXXXXXXXX101", "XXXX",
25 "XXXX", "XXXXXXXXXXXX" at the end of the third memory
cycle.

During the fourth memory cycle, the bit " $A_8 = 0$ " is
transferred from the memory to the memory buffer M0
during the first half of the memory clock cycle, and the
30 first PE shifts the contents of the data register YS0
left by one bit position and loads the vacated bit

-13-

position with the contents of the memory buffer M0 during the second half of the memory clock cycle. Fig. 3 therefore depicts the contents of the memory buffer M0 as "0", and the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXXXX1010", "XXXX", "XXXX", "XXXXXXXXXXXX" at the end of the fourth memory cycle. During the fifth memory cycle, the bit "A₇ = 0" is transferred from the memory to the memory buffer M0, and the data stored in the four (4) LSB positions of the data register Y0 are loaded into the communication register NEWS00 during the first half of the memory clock cycle; and, the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0 during the second half of the memory clock cycle. Fig. 3 therefore depicts the contents of the memory buffer M0 as "0", and the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXXXX10100", "1010", "XXXX", "XXXXXXXXXXXX" at the end of the fifth memory cycle. During the sixth memory cycle, the bit "A₆ = 1" is transferred from the memory to the memory buffer M0 during the first half of the memory clock cycle, and the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0 during the second half of the memory clock cycle. Fig. 3 therefore depicts the contents of the memory buffer M0 as "1", and the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXX101001", "1010", "XXXX", "XXXXXXXXXXXX" at the end of the sixth memory cycle.

-14-

During the seventh memory cycle, the bit "A₅ = 0" is transferred from the memory to the memory buffer M0 during the first half of the memory clock cycle, and the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0 during the second half of the memory clock cycle. Fig. 3 therefore depicts the contents of the memory buffer M0 as "0", and the respective contents of the YS0, NEWS00, NEWS11, and YS1 registers as "XXXXX1010010", "1010", "XXXX", "XXXXXXXXXXXX" at the end of the seventh memory cycle. During the first half of the eighth memory cycle, the bit "A₄ = 1" is transferred from the memory to the memory buffer M0, and the data stored in the communication register NEWS00 is loaded into the communication register NEWS11. During the second half of the eighth memory cycle, the first PE shifts the contents of the data register YS0 left by one bit position, and loads the vacated bit position with the contents of the memory buffer M0. Fig. 3 therefore depicts the contents of the memory buffer M0 as "1", and the respective contents of the YS0, NEWS00, NEWS11, and YS1 registers as "XXXXX10100101", "1010", "1010", "XXXXXXXXXXXX" at the end of the eighth memory cycle. During the first half of the ninth memory cycle, the bit "A₃ = 0" is transferred from the memory to the memory buffer M0, the data stored in the four (4) LSB positions of the data register Y0 are loaded into the communication register NEWS00, and the second PE loads the least significant four (4) bits of the data register YS1 with the contents of the communication register NEWS11. During the second half of

-15-

the ninth memory cycle, the second PE shifts the contents of the data register YS1 left by one bit position, and the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0. Fig. 3 therefore depicts the contents of the memory buffer M0 as "0", and the respective contents of the YS0, NEWS00, NEWS11, and YS1 registers as "XXX101001010", "0101", "1010", "XXXXXXX1010X" at the end of the ninth memory cycle.

During the first half of the tenth memory cycle, the bit " $A_2 = 0$ " is transferred from the memory to the memory buffer M0. During the second half of the tenth memory cycle, the second PE shifts the contents of the data register YS1 left by one bit position, and the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0. Fig. 3 therefore depicts the contents of the memory buffer M0 as "0", and the respective contents of the YS0, NEWS00, NEWS11, and YS1 registers as "XX1010010100", "0101", "1010", "XXXXXX1010XX" at the end of the tenth memory cycle. During the first half of the eleventh memory cycle, the bit " $A_1 = 1$ " is transferred from the memory to the memory buffer M0. During the second half of the eleventh memory cycle, the second PE shifts the contents of the data register YS1 left by one bit position, and the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0. Fig. 3 therefore depicts the contents of the memory buffer M0 as "1", and

-16-

the respective contents of the YS0, NEWS00, NEWS11, and YS1 registers as "X10100101001", "0101", "1010", "XXXXX1010XXX" at the end of the eleventh memory cycle. During the first half of the twelfth memory cycle, the
5 bit "A₀ = 1" is transferred from the memory to the memory buffer M0, and the data stored in the communication register NEWS00 is loaded into the communication register NEWS11. During the second half of the twelfth memory cycle, the second PE shifts the contents of the data
10 register YS1 left by one bit position, and the first PE shifts the contents of the data register YS0 left by one bit position and loads the vacated bit position with the contents of the memory buffer M0. Fig. 3 therefore depicts the contents of the memory buffer M0 as "1", and
15 the respective contents of the YS0, NEWS00, NEWS11, and YS1 registers as "101001010011", "0101", "0101", "XXXX1010XXXX" at the end of the twelfth memory cycle. At this time, it is noted that the YS0 register of the first PE contains the 12-bit operand "A".
20 During the first half of the thirteenth memory cycle, the data stored in the four (4) LSB positions of the data register Y0 are loaded into the communication register NEWS00, and the second PE loads the least significant four (4) bits of the data register YS1 with
25 the contents of the communication register NEWS11. During the second half of the thirteenth memory cycle, the second PE shifts the contents of the data register YS1 left by one bit position. It is noted that because every bit of the operand "A" has been loaded into the
30 data register YS0 during the first through twelfth memory cycles, the steps of shifting and loading the data

-17-

register YS0 no longer need to be performed during the second half of the memory cycle. Fig. 3 therefore depicts the contents of the memory buffer M0 as "X", and the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "101001010011", "0011", "0101", "XXX10100101X" at the end of the thirteenth memory cycle. During the fourteenth memory cycle, the second PE shifts the contents of the data register YS1 left by one bit position. Fig. 3 therefore depicts the contents of the memory buffer M0 as "X", and the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXXXXXXXX", "0011", "0101", "XX10100101XX" at the end of the fourteenth memory cycle. During the fifteenth memory cycle, the second PE shifts the contents of the data register YS1 left by one bit position. Fig. 3 therefore depicts the contents of the memory buffer M0 as "X", and the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXXXXXXXX", "0011", "0101", "X10100101XXX" at the end of the fifteenth memory cycle.

During the sixteenth memory cycle, the data stored in the communication register NEWS00 is loaded into the communication register NEWSI1, and the second PE shifts the contents of the data register YS1 left by one bit position. Fig. 3 therefore depicts the contents of the memory buffer M0 as "X"; and, the respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXXXXXXXX", "0011", "0011", and "10100101XXXX" at the end of the sixteenth memory cycle. Finally, during the seventeenth memory cycle, the data stored in the communication register NEWSI1 is loaded into the four (4)

-18-

LSB positions of the data register YS1, thereby completing the transfer of the 12-bit operand "A" from the first PE to the second PE. Fig. 3 therefore depicts the contents of the memory buffer M0 as "X", and the
5 respective contents of the YS0, NEWS00, NEWSI1, and YS1 registers as "XXXXXXXXXXXX", "0011", "0011", "101001010011" at the end of the seventeenth memory cycle. At this time, it is noted that the YS1 register of the second PE contains the 12-bit operand "A".

10 In this illustrative example, it should be understood that the step of loading in parallel the data stored in the communication register NEWS00 of the first PE into the communication register NEWSI1 of the second PE is performed concurrently with the process of serially
15 reading the operand "A" from the memory during the eighth and twelfth memory cycles, as described above. It should also be understood that the communication register NEWS00 of the first PE holds its data for nearly eight (8) PE clock cycles during the fifth through eighth memory
20 cycles, the ninth through twelfth memory cycles, and the thirteenth through sixteenth memory cycles to allow the data to settle before loading it into the communication register NEWSI1 of the second PE. It is also noted that the transfer of the 12-bit operand "A" from the first to
25 the second PE is completed in sixteen (17) memory cycles. In general, the SIMD array processor 100 transfers an n-bit operand between nearest neighboring PE's in $(n + 5)$ memory cycles. Such data transfers therefore have a
30 latency of five (5) memory cycles, as represented by, e.g., the first through fourth, and seventeenth memory cycles of the above-described illustrative example.

-19-

Having described the above embodiment, other alternative embodiments or variations may be made. Specifically, it was described that each of the PE's 104 through 134 of the SIMD array processor 100 (see Fig. 1) is connected to one and only one of the I/O's 0 through 15 of the RAM 102 to provide 1-bit serial access to data contained therein, and that the data register YS 204 of the PE 200 (see Fig. 2) is preferably 64-bits wide. However, it should be understood that the SIMD array processor may alternatively provide n-bit serial access to data stored in memory, and the YS register of the PE may alternatively be m-bits wide, wherein "m" is preferably greater than "n". The memory buffer M of the PE may therefore be n-bits wide to provide the n-bit serial data access. The SIMD array processor of this alternative embodiment provides enhanced data transfer efficiency by loading in parallel data stored in the communication register NEWS00 of a PE into the communication register NEWS11 of a neighboring PE while serially reading at least one (1) bit of an operand from memory.

Those of ordinary skill in the art should further appreciate that variations to and modification of the above-described SIMD array processor may be made without departing from the inventive concepts disclosed herein. Accordingly, the present invention should be viewed as limited solely by the scope and spirit of the appended claims.

-20-

CLAIMS

What is claimed is:

1. A single-instruction multiple-data array processor,
5 comprising:
at least one memory; and
a plurality of mesh-connected processing elements,
each processing element including at least one memory
buffer, at least one data register, at least one first
10 communication register, and at least one second
communication register, the memory buffer being adapted
to transfer data between the memory and the data
register, the first communication register being adapted
to transfer data between the data register and the second
15 communication register of a neighboring processing
element,
wherein each of the memory buffer, the data
register, and the first and second communication
registers has a width measured in bits, the width of the
20 data register being greater than the width of the memory
buffer and a multiple of the width of the memory buffer,
and the data register being at least as wide as the first
and second communication registers.
- 25 2. The single-instruction multiple-data array processor
of claim 1 wherein the first communication register is
adapted to transfer data to the second communication
register of the neighboring processing element while the
memory buffer transfers data between the memory and the
30 data register.

-21-

3. The single-instruction multiple-data array processor of claim 1 wherein each processing element is adapted to perform data processing operations using at least a portion of the data registers while the memory buffer transfers data between the memory and at least one of the data registers.

4. The single-instruction multiple-data array processor of claim 1 wherein the memory buffer is further adapted to transfer data serially between the memory and the data register.

5. The single-instruction multiple-data array processor of claim 1 wherein the first communication register is further adapted to transfer data in parallel between the data register and the second communication register of the neighboring processing element.

6. The single-instruction multiple-data array processor of claim 1 wherein the memory includes a plurality of bit-serial columns of memory locations, and the memory buffer is further adapted to transfer data serially between the data register and a selected one of the bit-serial columns of memory.

7. The single-instruction multiple-data array processor of claim 1 wherein the memory is a synchronous dynamic random access memory.

-22-

8. The single-instruction multiple-data array processor of claim 1 wherein the plurality of mesh-connected processing elements comprises a NEWS array.

5 9. A method of operating a single-instruction multiple array processor, the array processor including at least one memory and a plurality of mesh-connected processing elements, each processing element including a memory buffer, a data register, a first communication register,
10 and a second communication register, the method comprising the steps of:

transferring data between the memory and a first data register by a first memory buffer, the first data register and the first memory buffer being included in a
15 first processing element; and

transferring the data between the first data register and a second communication register by a first communication register, the first communication register being included in the first processing element and the
20 second communication register being included in a second processing element,

wherein a first portion of the data is transferred in the second transferring step while a second portion of the data is being transferred in the first transferring
25 step.

10. The method of claim 9 wherein the data are transferred in parallel between the first data register and the first communication register in the second
30 transferring step.

-23-

11. The method of claim 9 wherein the data are transferred between the memory and the first data register in the first transferring step in a bit-serial manner.

5

12. The method of claim 9 wherein the data comprise an n-bit data word and the data is transferred from the memory to the second communication register in the first and second transferring steps in $(n + 5)$ operating cycles of the memory.

10

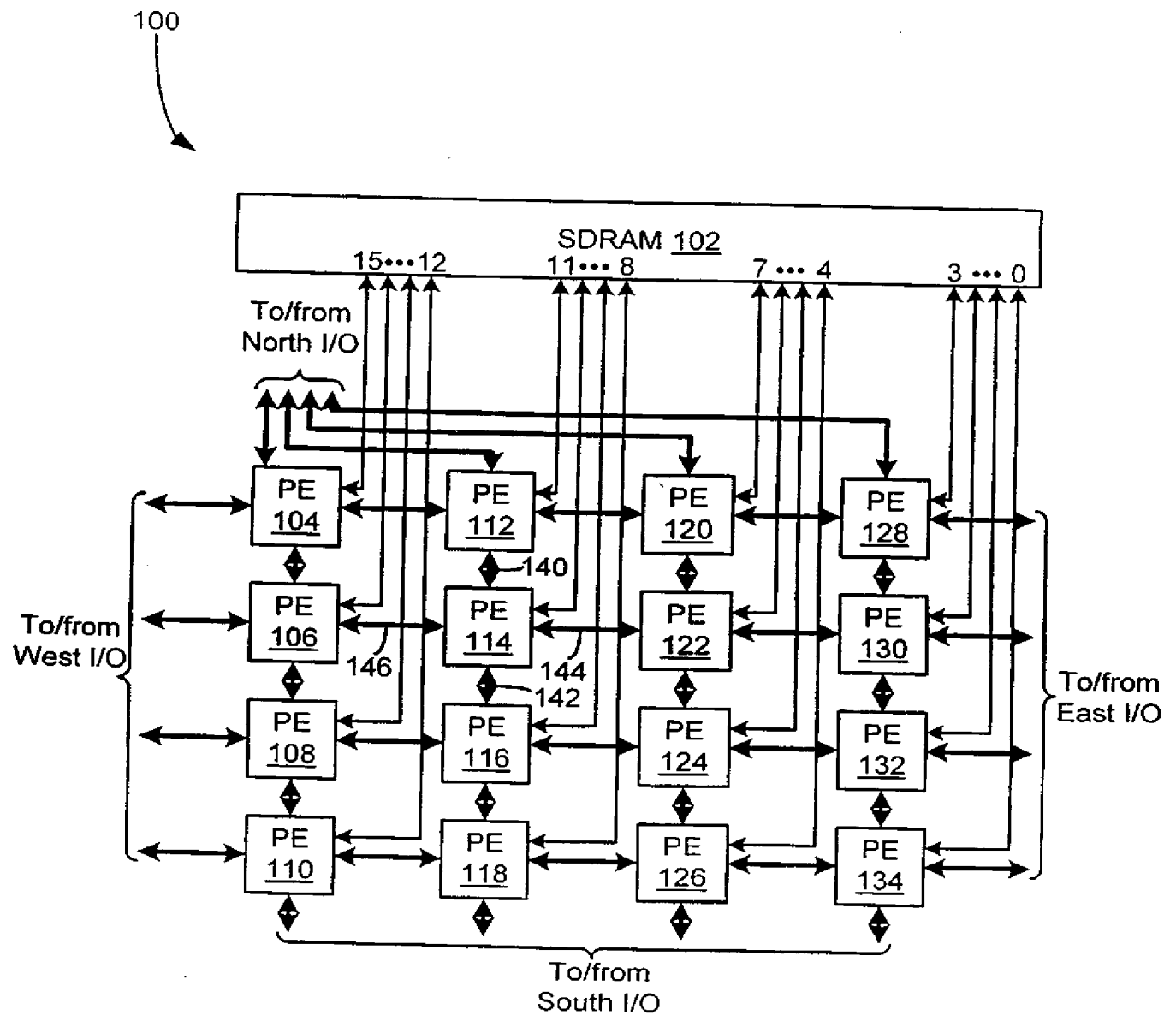


Fig. 1

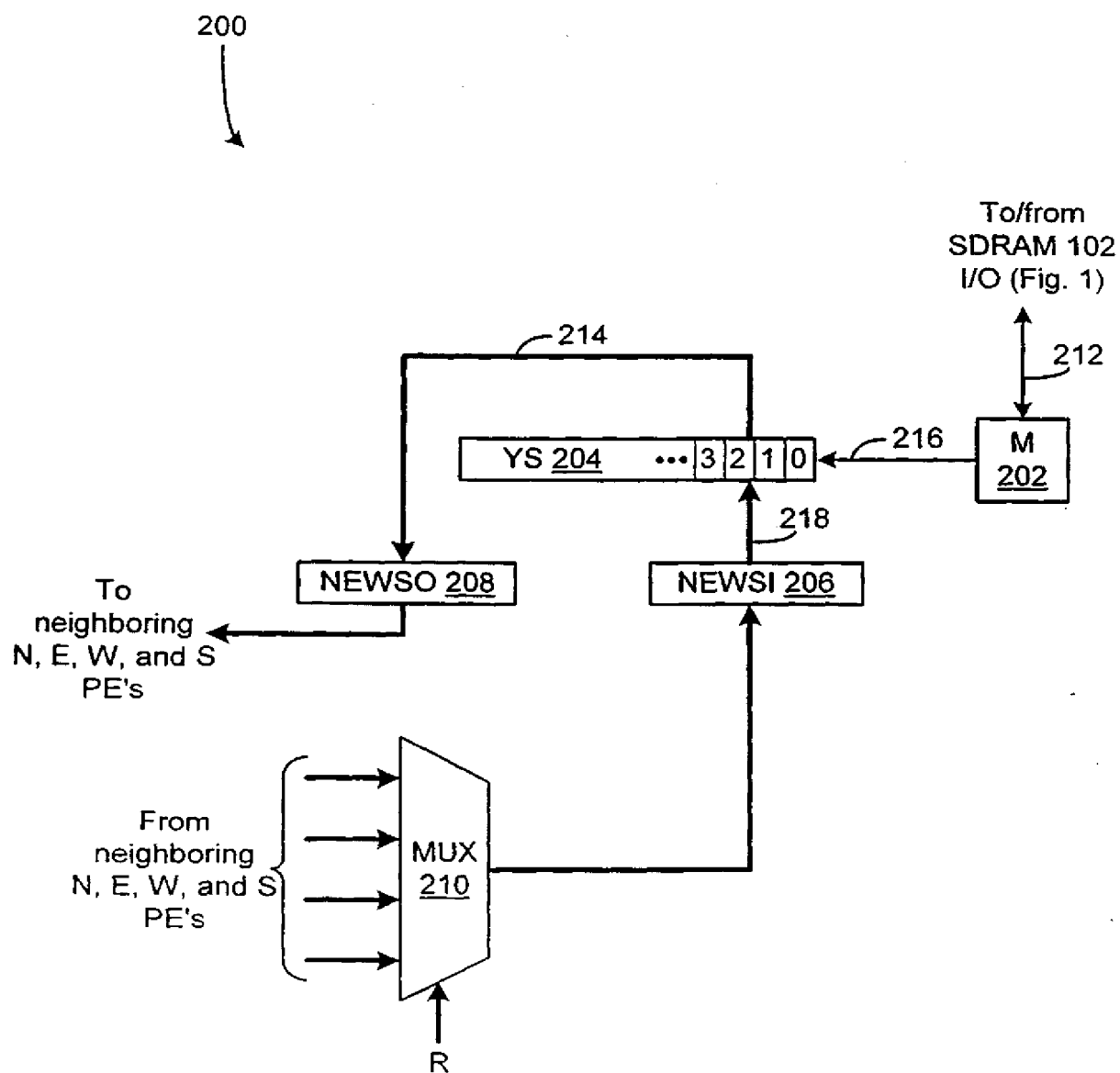


Fig. 2

Memory Cycle	M0	YS0 11...0	NEWS00 3...0	NEWS1 3...0	YS1 11...0
1	1	XXXXXXXXXXXX1	XXXX	XXXX	XXXXXXXXXXXX
2	0	XXXXXXXXXXXX10	XXXX	XXXX	XXXXXXXXXXXX
3	1	XXXXXXXXXXXX101	XXXX	XXXX	XXXXXXXXXXXX
4	0	XXXXXXXXXXXX1010	XXXX	XXXX	XXXXXXXXXXXX
5	0	XXXXXXXX10100	1010	XXXX	XXXXXXXXXXXX
6	1	XXXXXXX101001	1010	XXXX	XXXXXXXXXXXX
7	0	XXXXXX1010010	1010	XXXX	XXXXXXXXXXXX
8	1	XXXX10100101	1010	1010	XXXXXXXXXXXX
9	0	XXX101001010	0101	1010	XXXXXXXX1010X
10	0	XX1010010100	0101	1010	XXXXXXXX1010XX
11	1	X10100101001	0101	1010	XXXXX1010XXX
12	1	101001010011	0101	0101	XXXX1010XXXX
13	X	101001010011	0011	0101	XXX10100101X
14	X	XXXXXXXXXXXX	0011	0101	XX10100101XX
15	X	XXXXXXXXXXXX	0011	0101	X10100101XXX
16	X	XXXXXXXXXXXX	0011	0011	10100101XXXX
17	X	XXXXXXXXXXXX	0011	0011	101001010011

Fig. 3

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/31418 A3

(51) International Patent Classification?: **G06F 15/80**

(21) International Application Number: **PCT/US00/41571**

(22) International Filing Date: **25 October 2000 (25.10.2000)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:
60/161,587 26 October 1999 (26.10.1999) US

(71) Applicant: **PYXSYS CORPORATION [US/US]; 142 Q North Road, Sudbury, MA 01776 (US).**

(72) Inventors: **JACKSON, James, H.; 119 Penwood Drive, Cary, NC 27511 (US). KRAUS, Thomas, D.; 480 Walnut Street, Newtonville, MA 02460 (US).**

(74) Agents: **LEBOVICI, Victor, B. et al.; Weingarten, Schurigin, Gagnebin & Hayes, LLP, Ten Post Office Square, Boston, MA 02109 (US).**

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

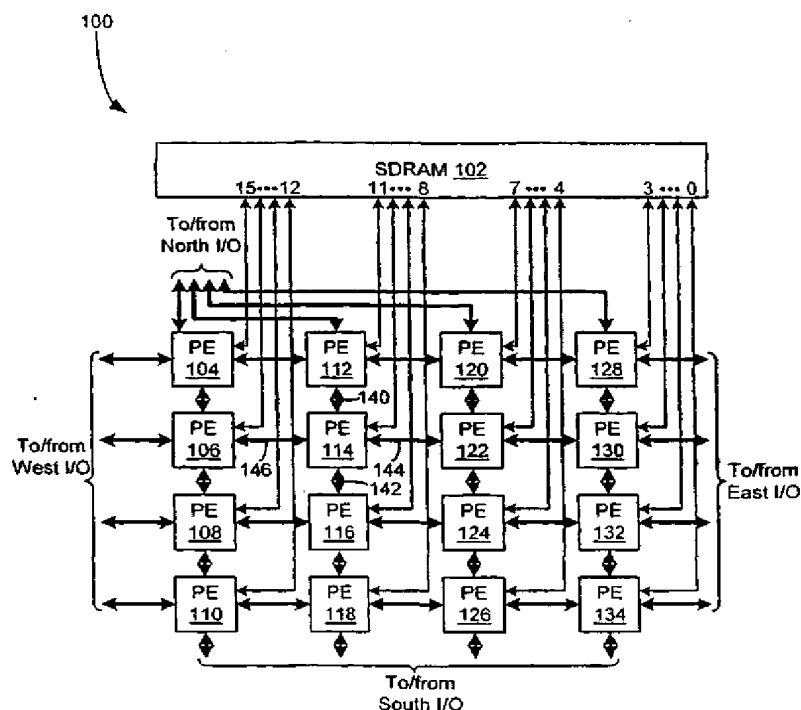
Published:

— with international search report

(88) Date of publication of the international search report:
6 December 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **WIDE CONNECTIONS FOR TRANSFERRING DATA BETWEEN PE'S OF AN N-DIMENSIONAL MESH-CONNECTED SIMD ARRAY WHILE TRANSFERRING OPERANDS FROM MEMORY**



(57) Abstract: A single-instruction multiple-data (SIMD) array processor (fig. 1) providing enhanced data transfer efficiency. The SIMD array processor includes at least one memory and a plurality of mesh-connected processing elements configured in an array. Each processing element (fig. 1, units 104, 106 etc.) in the array includes at least one "narrow" memory buffer, at least one "wide" data register, and at least one "wide" communication register. The narrow memory buffer is adapted to transfer data serially between the memory and the wide data register, the wide data register is adapted to transfer data directly to the wide communication register, and the wide communication register is adapted to transfer data directly to the communication register of a neighboring processing element while the memory buffer is accessing data from the memory.

WO 01/31418 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/41571

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 15/80

US CL : 712/22; 711/1, 101

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 712/22; 711/1, 101

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
IEEEElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN/CAS**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,933,650 A (VAN HOOK et al.) 3, August 1999, abstract, col. 1, lines 25-56; col. 4, lines 34-64 ; col. 5, lines 8-65	1-12
Y	US 5,164,913 A (GUILFOYLE et al.) 17, NOVEMBER, 1992, col. 42, lines 18-40	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

09 APRIL 2001

Date of mailing of the international search report

26 APR 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

Eddie P. Chan

Telephone No. (703) 305-9712

Form PCT/ISA/210 (second sheet) (July 1998) *